Chapter 7

Conclusion and Future Work

Action recognition based on videos has become a common and effective approach to understand human behaviors. Compared to image-based action recognition, videos provide more information that can help reduce ambiguity and provide context for the actions being performed. In recent years, advances in datasets, models, and learning approaches have significantly improved the accuracy and effectiveness of video action recognition. However, there are still challenges and unsolved problems in action recognition such as representation bias, human behavior traits, fall, unusual, and multi-label action.

Conclusion

The focus of this thesis is human action recognition, which is a critical area of application in computer vision. Through this research, the aim is to develop algorithms and models that can accurately detect and classify human actions from visual data, such as images or videos. Its primary aim is to accurately describe human actions in presence of representation bias, human behaviour traits, action in wild, and multi-label complex action. Human action recognition has a wide range of applications, including video surveillance, healthcare, robotics, human-computer interaction, and sports analytics. While progress has been made in this area, challenges remain due to factors

such as variability in representation bias, human behaviour traits, action in wild, and multi-label action. The computer vision community has focused particularly on deep learning in recent years. This research provides a summary of state-of-the-art action recognition using deep learning and video analysis. The most significant deep learning models for human action recognition are presented, and their strengths and weaknesses are highlighted through analysis to show the current state of deep learning algorithms used to address human action recognition issues in realistic videos. Our study identifies state-of-the-art deep architectures in action recognition based on the quantitative analysis using recognition accuracy provided in the literature and then presents current trends and open difficulties for subsequent works in this field.

Future Directions

In the future, this work can be extended and can be used to solve the following problems:

- Working with untrimmed videos in the wild. Untrimmed video are longer in duration with multiple tasks. It requires continuous recognition over multiple actions. This requires training of a large video with high computational complexity.
- Working for distinguishing actions that are part of imitation or acting from actual events. A long duration video will have a longer temporal complex, which will help to distinguish a original one with a imitated one.